

CuGene as a tool to view and explore genomic data

authors:

**Michał Haponiuk, Magdalena Pawełkiewicz,
Zbigniew Przybecki, Robert M. Nowak**

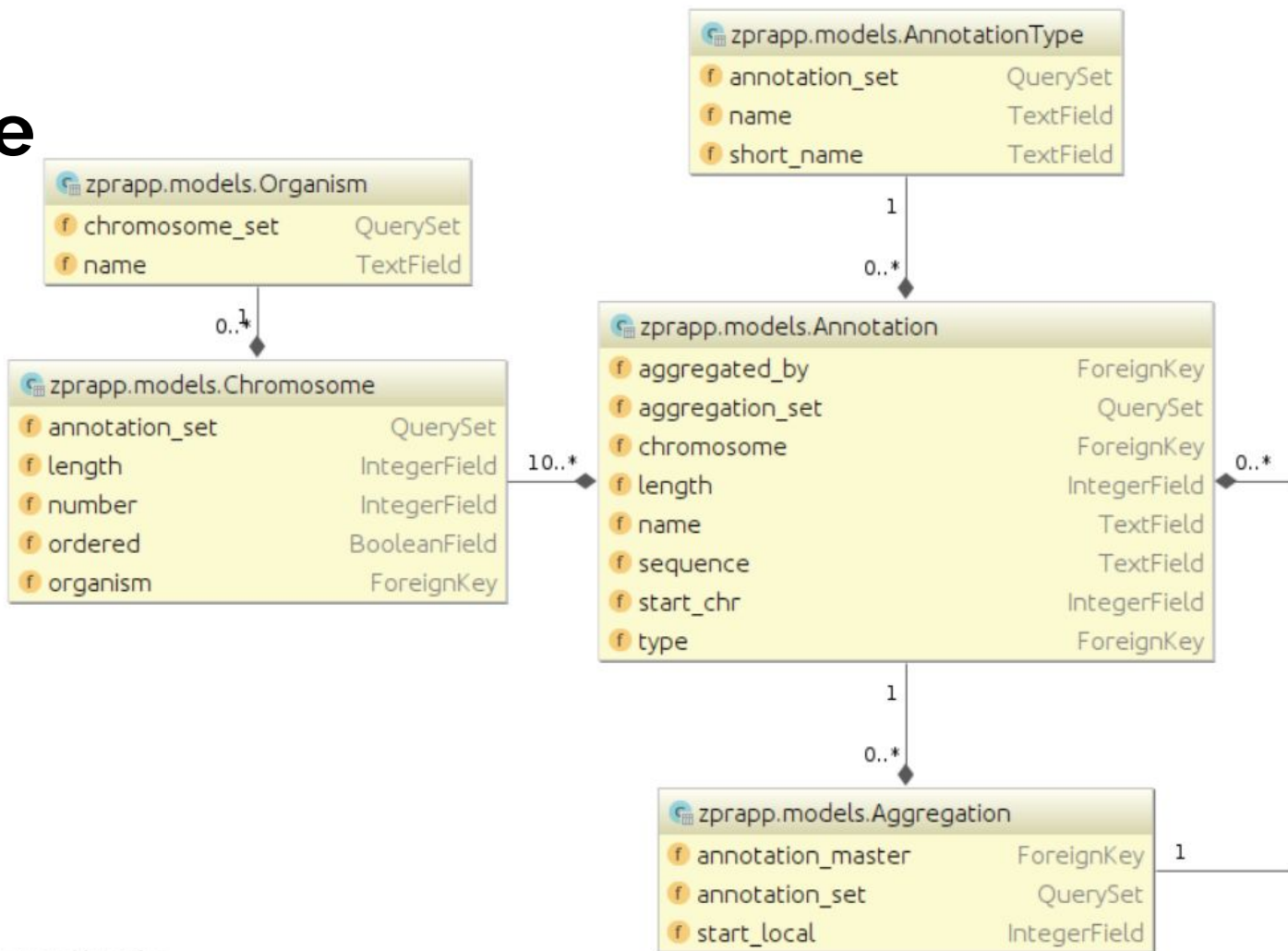
Software functionality

- display of sequence and annotation tracks
- interactive zooming
- editing of individual features, supporting manual annotation
- intuitive aligned sequences, elegant to view
- many searching algorithms
- import FASTA, GFF (general feature format), XLS (Excel) files
- the possibility to integrate third party applications with CuGene via HTTP API

Architecture



Database



Mapped cucumber's genome

Chromosome	chr. length [Mbp]	No. mapped contigs	Mapped length [bp]	Mapped length [%]
1	55	22	32970425	59.9%
2	44	13	23992470	52.2%
3	65	14	39532546	60.8%
4	61	15	24781874	40.6%
5	49	22	26573846	54.2%
6	42	19	29507078	70.2%
7	52	11	18951584	36.4%
Σ	368	116	196309823	53.3%

- All contigs amount: 8035
- Mapped contigs amount: 116
- Sum of known contig sequence lengths:
342 288 160 bp
- Sum of mapped contig sequence lengths:
196 309 823 bp (57%)

Chen, J.-f., Staub, J. E., and Jiang, J., A reevaluation of karyotype in cucumber (*cucumis sativus* L.), Genetic Resources and Crop Evolution 45 (4), 301305 (1998).

Pattern searching algorithms

accurate match:

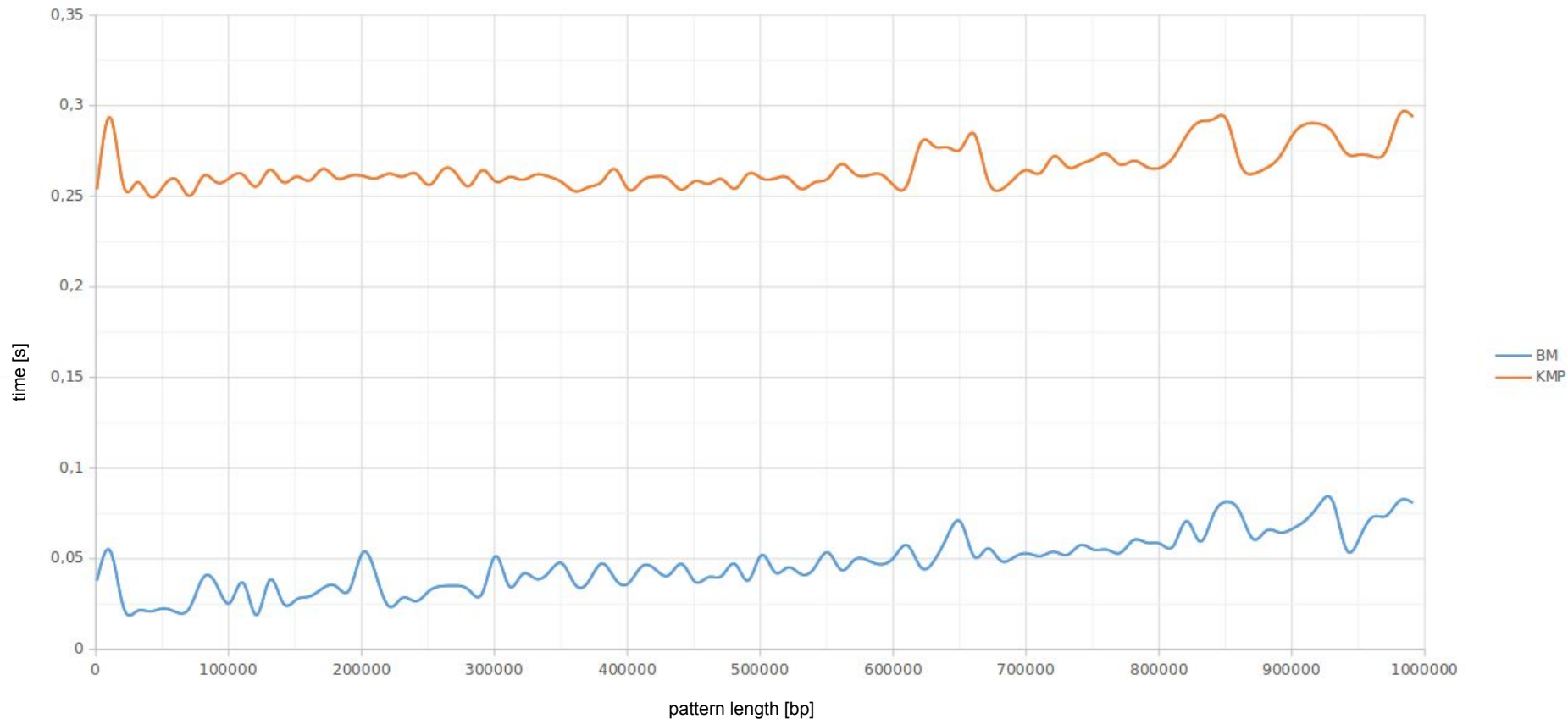
- Knuth-Morris-Prath
- Boyer Moore

approximate match:

- Smith-Waterman (dynamic programming)
- BLAST (heuristic method)

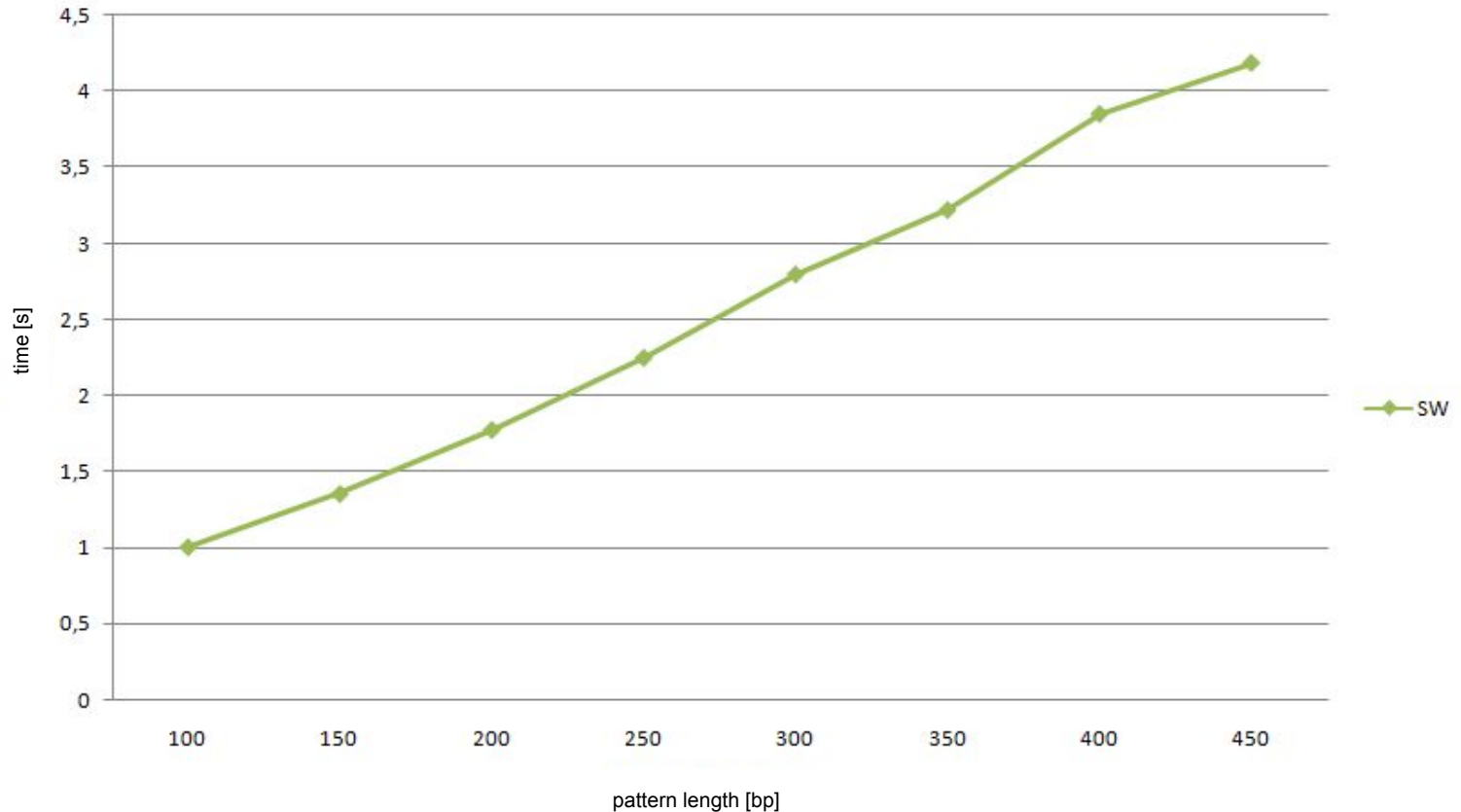
Searching time, KMP, BM

sequence length: 10 000 000 bp



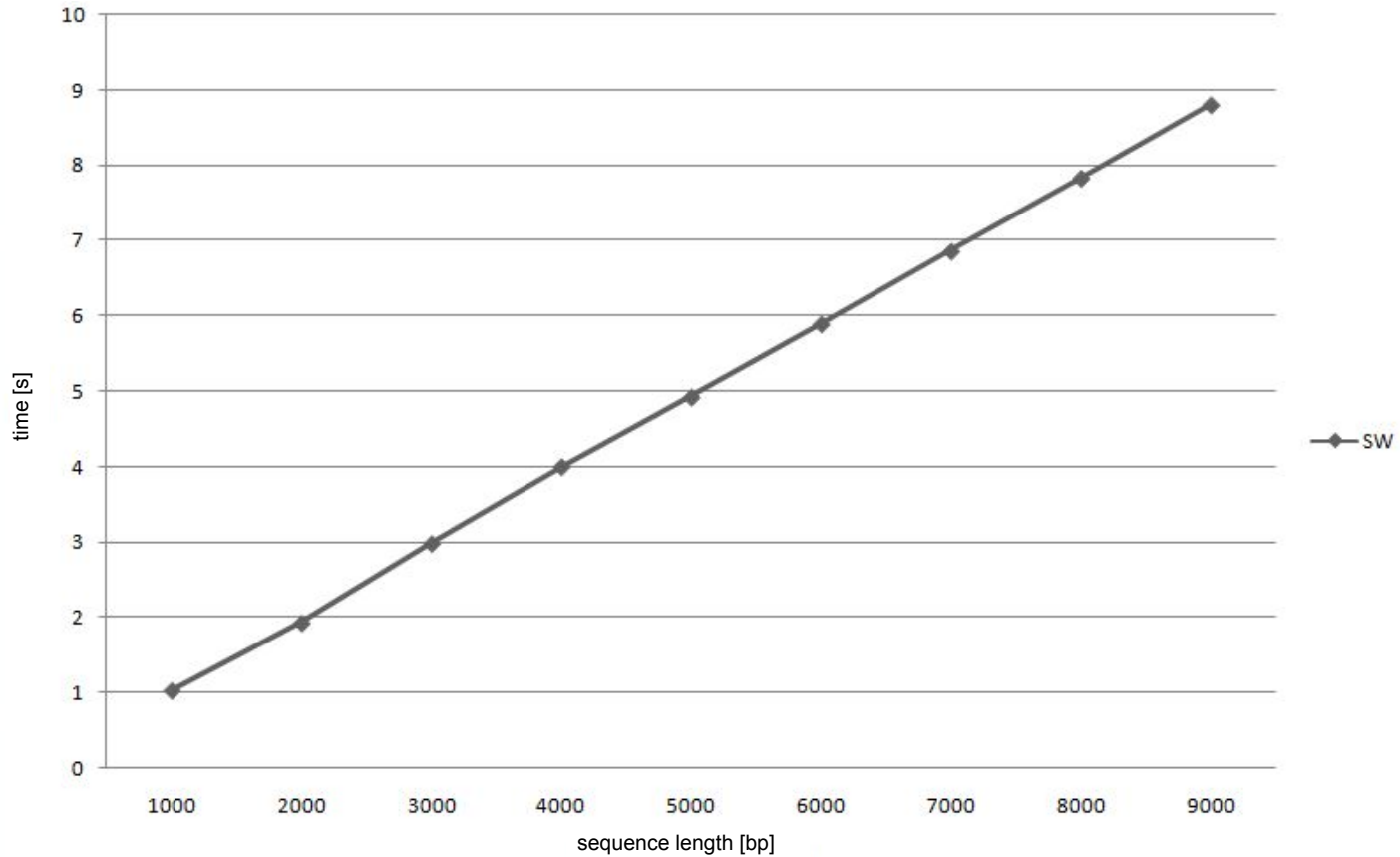
Searching time, SW algorithm

sequence length: 5 000 bp



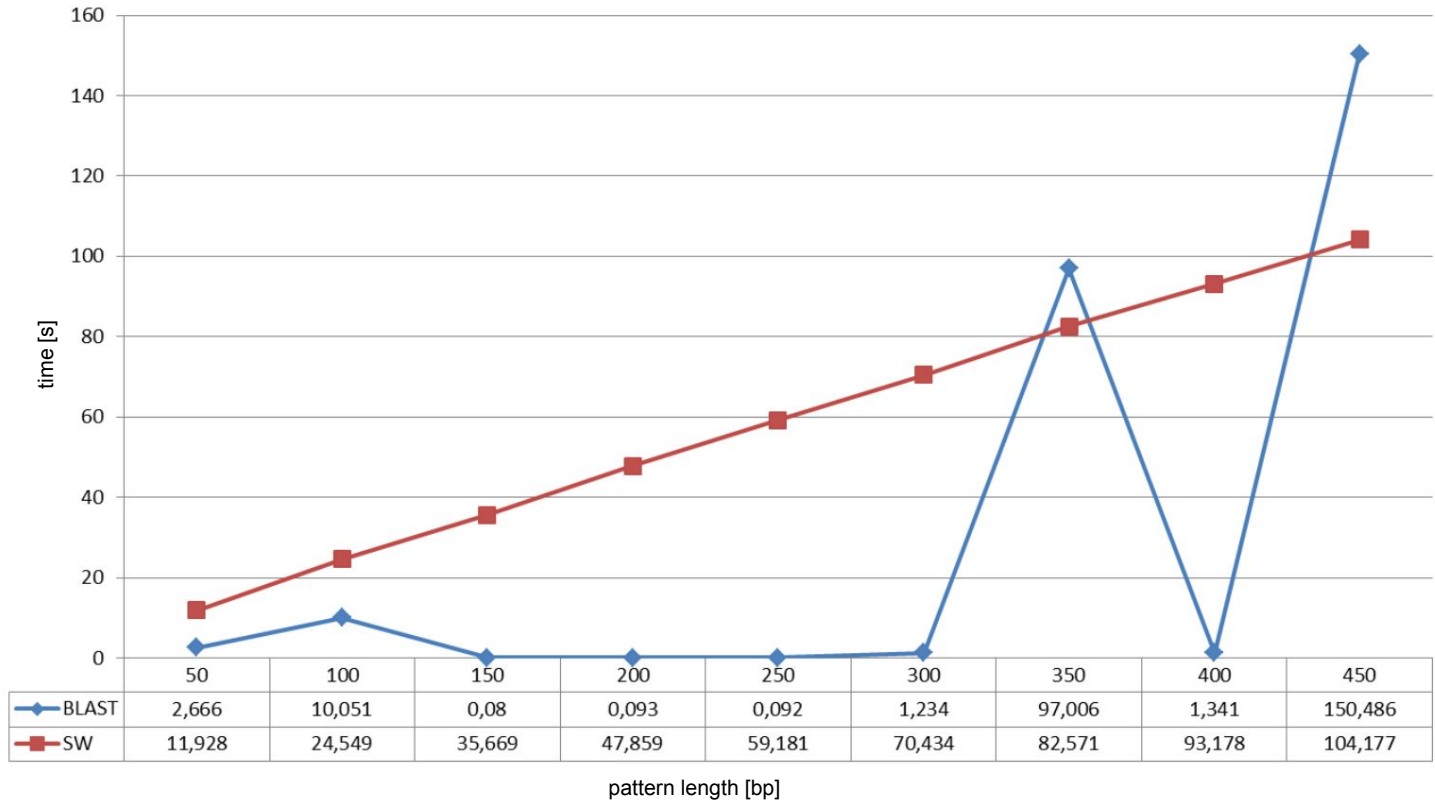
Searching time, SW algorithm

pattern length: 500 bp



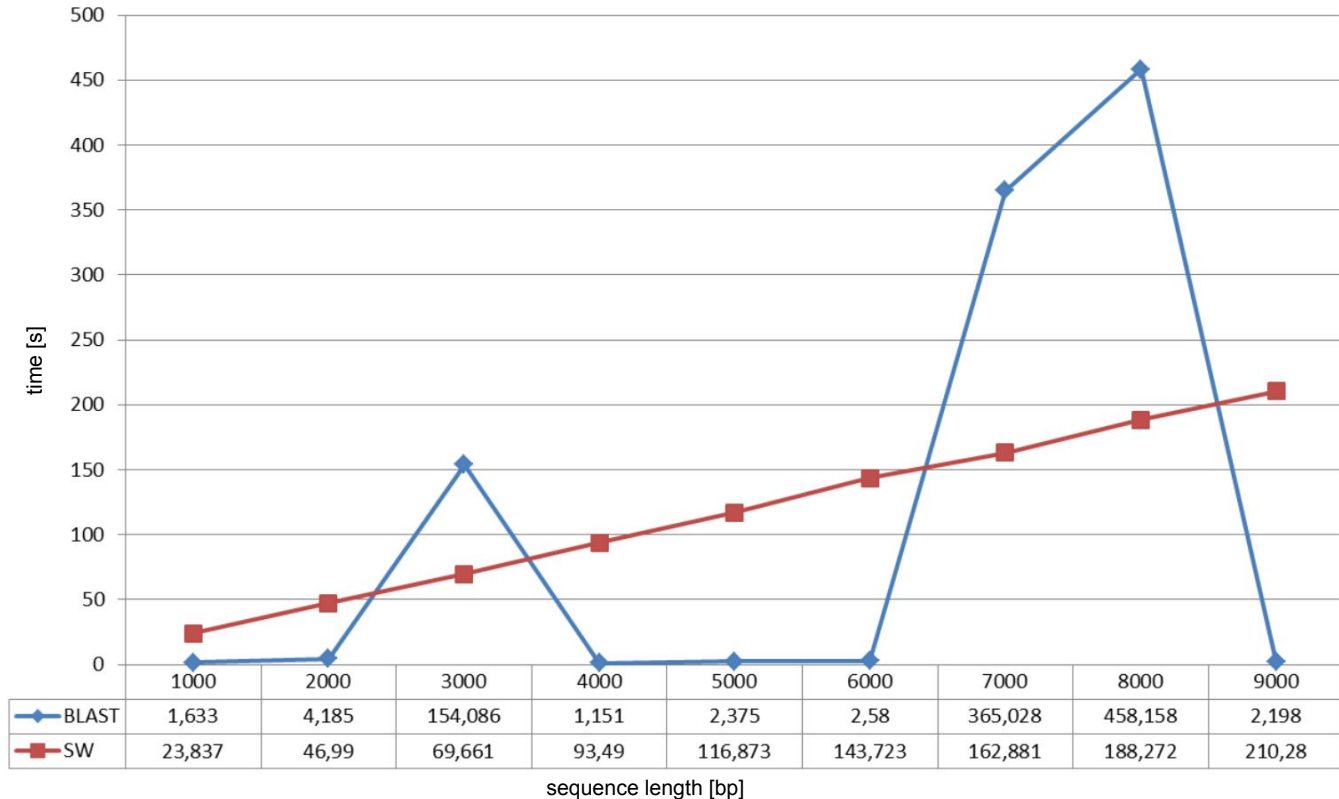
Searching time, SW and BLAST

sequence length: 5 000 bp



Searching time, SW and BLAST

pattern length: 500 bp



Genome browsers functionality comparison

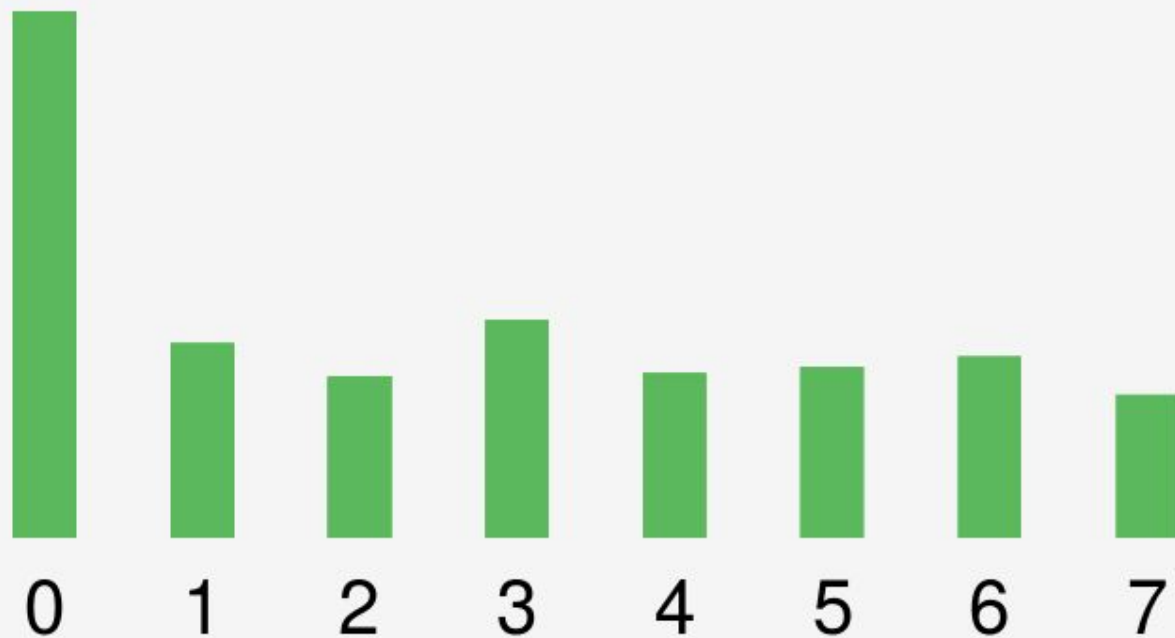
	Embedded data	Custom species		Searching			Frame-work	Stand-alone	Mobile devices
		one	many	BLAST	SW	KMP			
Ensembl	+	+	+	+	-	-	+	+	+
UCSC	+	+	+	+	-	-	+	+	+
Gramene	+	+	+	+	-	-	-	-	+
Gbrowse	-	+	+	+	-	-	+	+	+
Jbrowse	-	+	+	+	-	-	+	+	+
Abrowse	-	+	+	+	-	-	+	+	+
Anno-J	-	+	+	+	-	-	+	+	+
SynBrowse	-	+	+	-	-	-	+	+	-
SynView	-	+	+	-	-	-	+	+	+
LookSeq	-	+	+	+	-	-	+	+	+
Vista	+	-	-	+	-	-	-	-	-
Celera Genome Browser	-	+	-	-	ND	ND	+	+	-
Alamut	+*	-	-	+	ND	ND	-	-	-
GeneWall	+*	-	-	-	-	-	-	+	+
GenomeView	-	+	+	+	-	-	+	+	-
CuGene	-	+	+	+	+	+	+	+	+

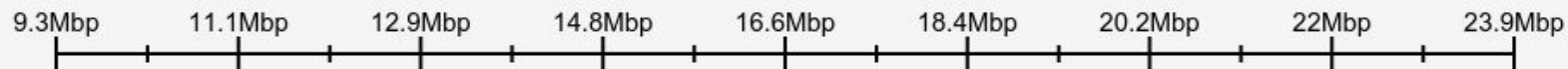
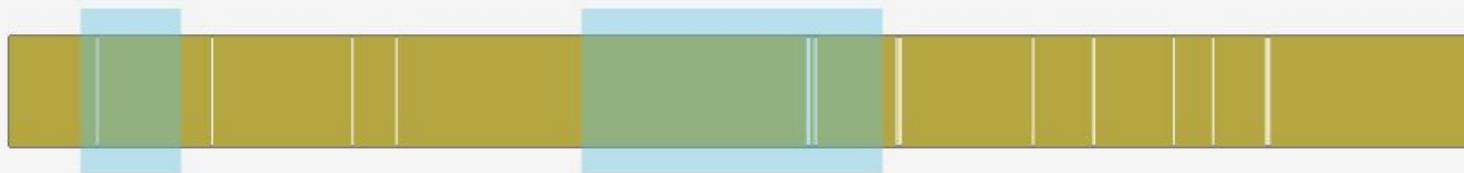


CuGene



0





Search: %

od: do:

☒ MARKERY

← 14.58Mbps →

Conclusions

- client-server web application, easy to use
- developed to store many organisms, many genomes
- used for cucumber genome
- freely available at <http://cugene.sourceforge.net>
- demo server <http://antakya.ise.pw.edu.pl:9002>

Thank you